Physics 4460/5460
Week Se7en
- what happened?

Day 12:

Assessment (Evaluation of Instruction – Redish)

Class-updates:

- We'll continue to update schedule
- Keep suggestions coming.. We'll survey next week

Learning Goals:

- be able to select among the alphabet soup of Conceptual Surveys
- interpret/ apply results
- describe the design / validation
- select other tools for other evaluation goals

Next Steps

Given the following history and prospects of topics, what would you like to see next?

So Far

- State of Affairs
- Theory: constructivism
- Theory: socio-cultural
- (Mis)conceptions
- Attitudes/ Beliefs
- Assessment
- Metacognition
- Goals

Possible

- Problem Solving
- Diversity / Inclusion
- Sustaining/ Scaling
- Representational formats
- Analogy
- Historical Perspectives
- Laboratories
- Politics of education
- More content areas …
- …

Warm-Up

Reflect on a session / course / teaching experience that was really good.

How do you know that the experience was good?

what evidence do you use to backup the claims of effectiveness?

Example… consider finals

For typical Physics 1 course are Grades and the Final Exam effective forms of evaluation of the course?

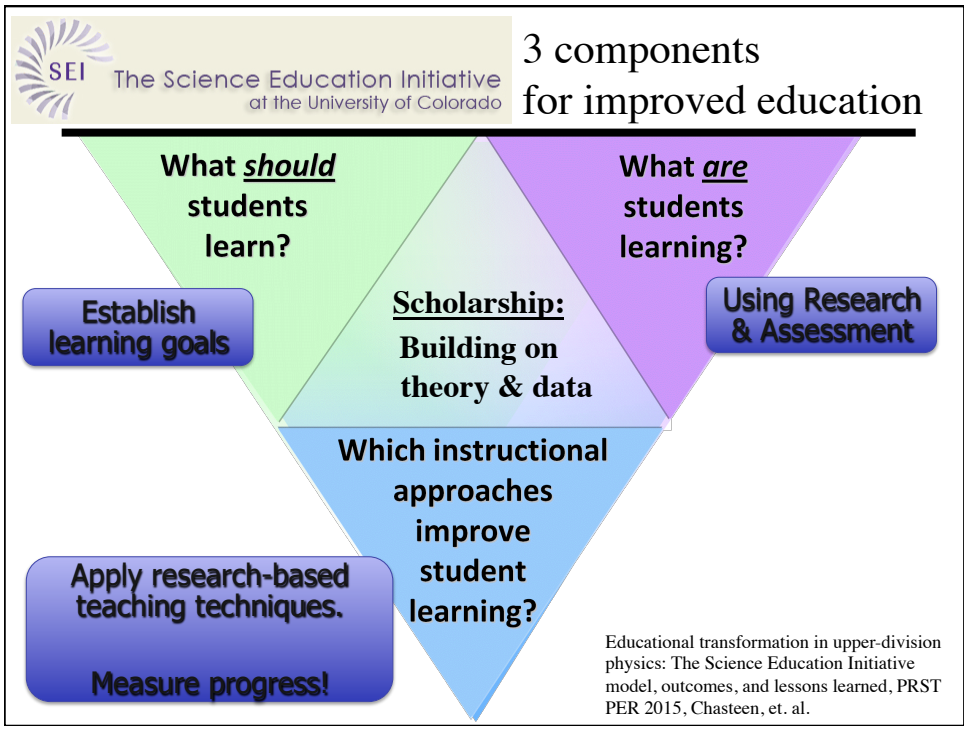
I. Yes

II. No

III. it depends

PI example
Indiv. then Group

- **Actually talking about and agreeing on these goals upfront might be useful…**



Assessment is about Feedback (and acting on that feedback)

to *whom*?

for *what* reason?

& *when* ?

Dimensions of assessment

What: Content/process ←→ Affect/perception

When/why: Formative ←→ summative

Where: Individual concept ←→ Entire course

How: Research based ←→ informal

Who: Students ←→ Faculty (Department)

What are our goals in class?

Novice		Expert
Formulas & "plug 'n chug"	content	Concepts & Problem Solving
Pieces	structure	Coherence
By Authority	process	Independent (experiment)
Drudgery	affect	Joy

think about science like a scientist
think about education like a scientist

Adapted from: Hammer (1997) COGNITION AND INSTRUCTION (physics),

Learning Goals.. specifically

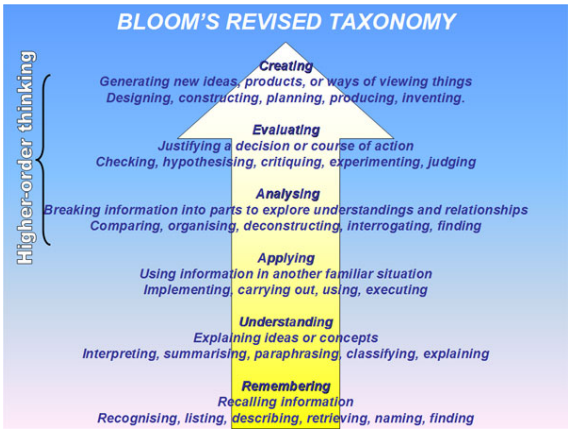
- Look at Goals?
- If we don't decide others will.

E.g. E/M Learning Goals

COURSE SCALE LEARNING GOALS E&M 1

- **Math/physics connection:** Students should be able to
- Visualize **the problem:** Students should be able to
- **Organized knowledge:** Students should be able to
- **Communication**
- **Problem-solving techniques:**
  - **5a. Approximations:**
  - **5b. Series expansions: …**
- Problem-solving **strategy:**
- **Expecting and checking solution:**
- **Intellectual maturity:**
- **Maxwell's Equations.**
- **Build on Earlier Material.**

Blooms Taxonomy



For Thurs, expand on Bloom

The Knowledge Dimension	The Cognitive Process Dimension					
	1. Remember	2. Understand	3. Apply	4. Analyze	5. Evaluate	6. Create
A. Factual knowledge						
B. Conceptual knowledge						
C. Procedural knowledge						
D. Meta-cognitive knowledge						

??

Fine… those (and others) are our goals… how do we know if we're there

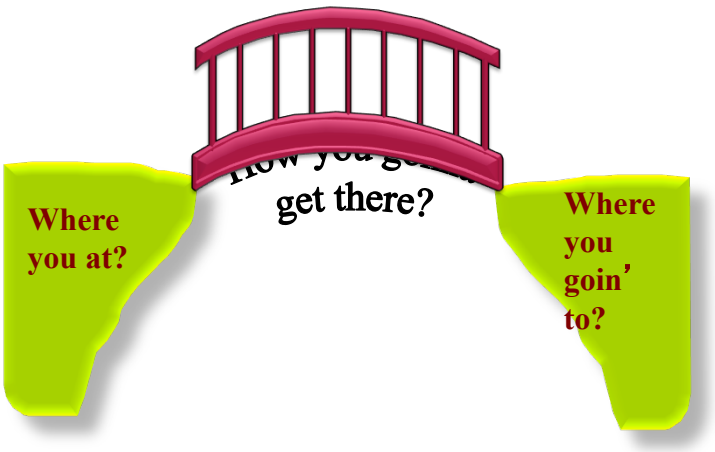
- Measurement
- Key distinction:
  - **Formative:** to provide feedback, be a mechanism for learning / change
  - **Summative:** after the fact to measure what happened.
- Key question:
  - When / who would want to use which type?
  - What actually happens?

??

Goals for Assessment

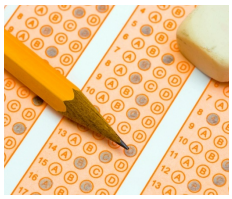
- Improve a learning opportunity
  - Concept test
  - develop skills of scientific practice (talking, justifying, arguing, logically deducing . . .)
- Improve a lecture period / unit
  - conceptual mastery
  - problem solving acuity
- Improve a course
- Improve the department
- Improve society?

The Formative Assessment Processes:



Atkin, Black, & Coffey 2001; Otero & Nathan 2008

Why Formative Assessment?



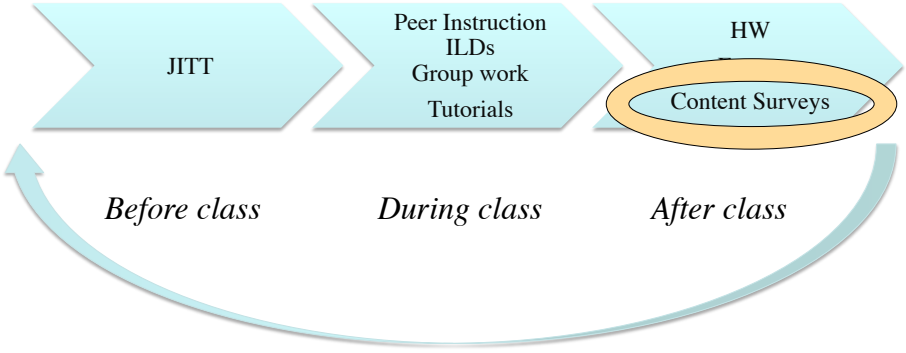
Improves Achievement

Shifts Classroom practices

Underpins Reform

Black & William 1998a, 1998b
Andreade & Cizek 2010

When to assess



Listening to Students

If I had to reduce all of educational psychology to just one principle, I would say this:

The most important single factor influencing learning is what the student already knows.

Ascertain this and teach him accordingly.

-D. P. Ausubel [Ausubel 1978 in Redish2003]

Assessment

- Seek to measure what we're after. How?
  - Pre/ Post-Tests
  - Interviews
  - Project-based work
  - Individual / Collaborative
  - Time bound / not-time bound

Focus on Content Mastery (mostly)

Validity and Reliability

Valid:

– are we measuring what we think we are?

– is the instrument internally consistent?

– would a physicist see this test as physics?

Reliable:

- will students give the same response on two subsequent rounds of inquiry (no teaching in between)?

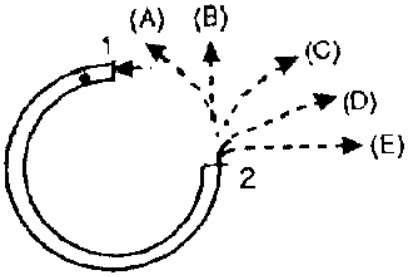
A possible "tipping" point

- **Force Concept Inventory***
- Multiple choice survey, (pre/post)
- Experts (especially skeptics!)

  necessary (not sufficient) indicator of conceptual understanding.

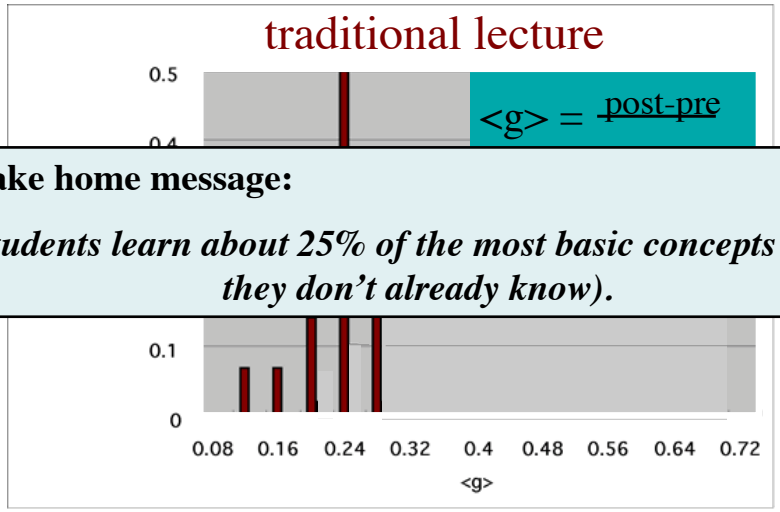
* Hestenes, Wells, Swackhamer, Physics Teacher 20, (92) 141

Sample question



Looking down at a track (flat on table), a ball enters at point 1 and exits at point 2. Which path does it follow as it exits (neglect all friction)?

Force Concept Inventory

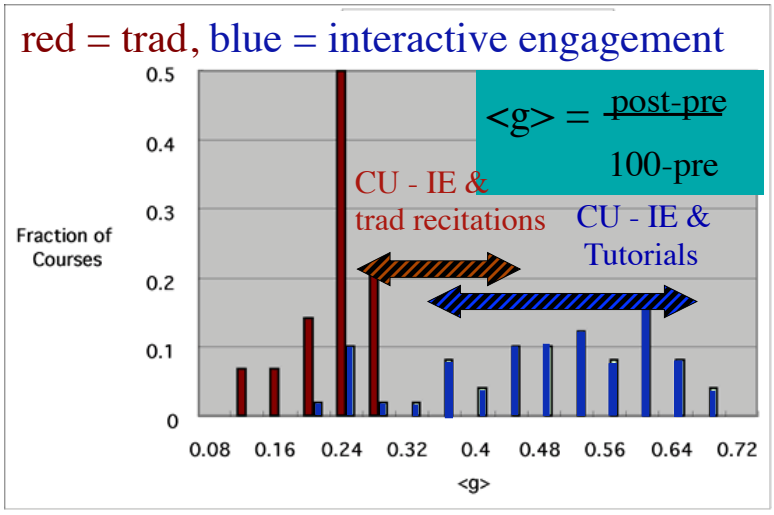


Take home message:

Students learn about 25% of the most basic concepts (that they don't already know).

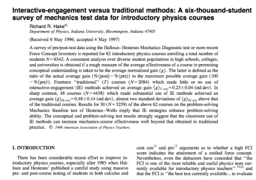
R. Hake, "…A six-thousand-student survey…" AJP **66**, 64-74 ('98).

Force Concept Inventory



R. Hake, "…A six-thousand-student survey…" AJP **66**, 64-74 ('98).
Pollock and Finkelstein, PRST PER (2008).

FCI. Out of Date?



Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses

Richard R. Hake[a]
Department of Physics, Indiana University, Bloomington, Indiana 47405

(Received 6 May 1996; accepted 4 May 1997)

A survey of pre/post-test data using the Halloun–Hestenes Mechanics Diagnostic test or more recent Force Concept Inventory is reported for 62 introductory physics courses enrolling a total number of students $N=6542$. A consistent analysis over diverse student populations in high schools, colleges, and universities is obtained if a rough measure of the average effectiveness of a course in promoting conceptual understanding is taken to be the average normalized gain $\langle g \rangle$. The latter is defined as the ratio of the actual average gain $(\%\langle post \rangle - \%\langle pre \rangle)$ to the maximum possible average gain $(100 - \%\langle pre \rangle)$. Fourteen "traditional" (T) courses ($N=2084$) which made little or no use of interactive-engagement (IE) methods achieved an average gain $\langle g \rangle_{T\text{-ave}} = 0.23 \pm 0.04$ (std dev). In sharp contrast, 48 courses ($N=4458$) which made substantial use of IE methods achieved an average gain $\langle g \rangle_{IE\text{-ave}} = 0.48 \pm 0.14$ (std dev), almost two standard deviations of $\langle g \rangle_{IE\text{-ave}}$ above that of the traditional courses. Results for 30 ($N=3259$) of the above 62 courses on the problem-solving Mechanics Baseline test of Hestenes–Wells imply that IE strategies enhance problem-solving ability. The conceptual and problem-solving test results strongly suggest that the classroom use of IE methods can increase mechanics-course effectiveness well beyond that obtained in traditional practice. © *1998 American Association of Physics Teachers.*

I. INTRODUCTION

There has been considerable recent effort to improve introductory physics courses, especially after 1985 when Halloun and Hestenes[1] published a careful study using massive pre- and post-course testing of students in both calculus and cent con[11] and pro[12] arguments as to whether a high FCI score indicates the attainment of a unified force concept. Nevertheless, even the detractors have conceded that "the FCI is one of the most reliable and useful physics tests currently available for introductory physics teachers" [11(a)] and that the FCI is "the best test currently available... to evaluate

Sunday (Oct 16, 2011), PhysLRNR: (Hake)

"Robin Millar and Jonathan Osborne in Chapter 3 of "Research and Practice: A Complex Relationship" [Shelley et al. (2009)] claimed that:

(a) NO STANDARD OR COMMONLY AGREED OUTCOME MEASURES EXIST FOR ANY MAJOR TOPIC IN SCIENCE EDUCATION. . . . [[my CAPS]]. . . ,

(b) the Force Concept Inventory (FCI) reflects a choice of *values* that is arguable, and

(c) the FCI has not been subjected to the same rigorous scrutiny of factorial structure and content validity as have standard measures in psychology"

How Scholarship happens?

Hake, R.R. 2011. "No Standard Outcome Measures For Science Education? #2" online on the OPEN! AERA-L archives at <http://bit.ly/rfyamc>..

http://listserv.aera.net/scripts/wa.exe?A2=AERA-L;e843603a.1110

"50 years of research, curriculum development, and implementation have not presented consistent and compelling patterns of outcomes."

Shelley et al. (2009, p. 4) summarizing a claim by Osborne (2007)

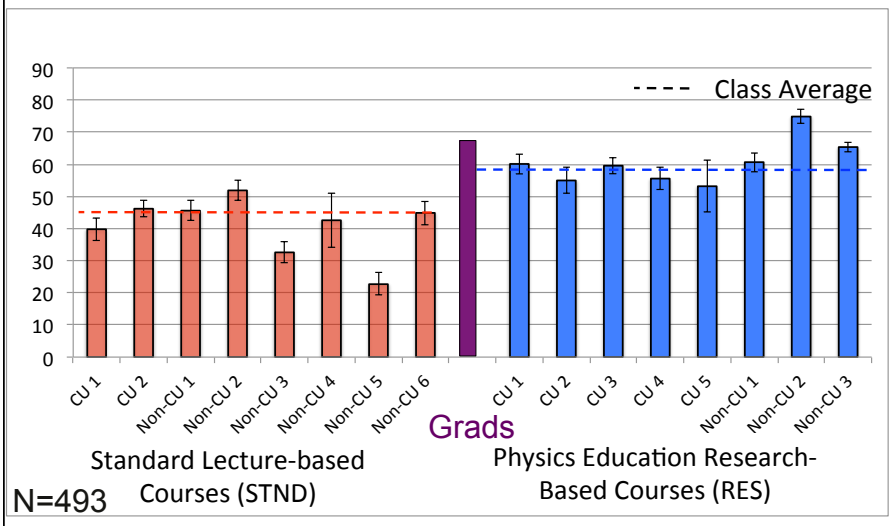
"Physics educators have led the way in developing and using objective tests to compare student learning gains in different types of courses, and chemists, biologists, and others are now developing similar instruments. These tests provide convincing evidence that students assimilate new knowledge more effectively in courses including active, inquiry-based, and collaborative learning, assisted by information technology, than in traditional courses."

Wood & Gentile (2003) Science "Teaching in a research context"

Is the FCI / FMCE it?

- Let's look at CUE / QMAT

CUE Results: Comparison



Chasteen et al, PERC 2011, submitted to AJP

Resources and Guides for Use/Interpretation

I. At the lower division included are:
- The Force and Motion Conceptual Evaluation (FMCE)
- The Force Concept Inventory (FCI)
- The Brief Electricity and Magnetism Survey (BEMA)
- The Conceptual Survey of Electricity and Magnetism (CSEM)

II. At the Upper division we include materials in development at CU
- Classical Mechanics (CCMI)
- Electrostatics (CUE)
- Electrodynamics (CURrENT)
- Quantum Mechanics (QMAT)

III. And two Beliefs Instruments:
- Colorado Learning Attitudes about Science Survey (CLASS)
- E-CLASS (version for experimental physics)

Many many more….

https://www.physport.org/assessments/

How Do I Use these?

What do I do with the results?

physport.org (much more coming soon)

Other ways to assess (content) mastery?

What types of assessment should you do?

Answering requires

- Clear identification of goals
- Consideration of what is measurable

Categories of Assessments: Assignments and exams♪

- Rubrics
  - Specify performance criteria
  - Help students see learning goals; guide efforts
  - Guide instructor grading
- Scoring codes
  - More feedback to students (but more generic)
  - Evaluate frequency of different approaches/errors
- New Models of exams
  - Two-stage exams
  - Standards based
  - Practicing what we teach

Example rubric

	Ability to design and conduct a testing experiment			
Scientific ability	0 (Missing)	1 (Inadequate)	2 (Needs some improvement)	3 (Adequate)
Is able to design a reliable experiment that tests the relationship or explanation.	The experiment does not test the relationship or explanation.	The experiment tests the relationship or explanation, but due to the nature of the design it is likely the data will lead to an incorrect judgment.	The experiment tests the relationship or explanation, but due to the nature of the design there is a moderate chance the data will lead to an inconclusive judgment.	The experiment tests the relationship or explanation and had a high likelihood of producing data that will lead to a conclusive judgment.

Etkina, et al., PRST-PER **2**, 020103 (2006)

Example scoring code

Problem 5
This problem can be analyzed through conservation of momentum. The carts' initial momentum is not zero. This is similar to example 9.3, except that the carts do not start at rest. Problem solving strategy 9.1 on p251 is useful here.

Code	Description
G	Good job
H	Used conservation of momentum approach, but made a minor calculation error, most commonly a sign error with the final momentum of the light cart (remember, it is going opposite to it's initial direction, you have to represent this in the math).
I	Used conservation of momentum approach, but made a physics error in the solution. The most common error was not including the two cart's initial momentum (remember, they are moving to the right when the spring goes off).
J	Tried to solve the problem with conservation of kinetic energy. While mechanical energy *is* conserved in this situation, you must include the spring's potential energy, which is converted to kinetic energy. The carts' total kinetic energy increases by an amount equal to the PE stored in the compressed spring.
K	Tried to use the elastic collision equations 10.43 (p285). These equations are only valid when one of the objects is initially at rest, which is not the case in this situation.
L	Other partial attempts.
M	No work.

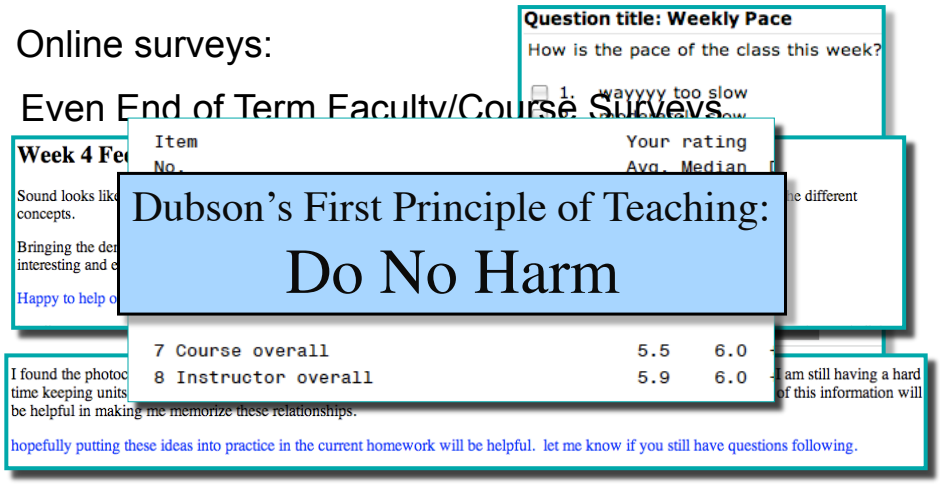
Two Stage Exam
UBC (see: http://www.cwsei.ubc.ca)

1st stage: individual (traditional) - ~66% of time
Turn in
2nd stage: collective response (same exam)
Final score is a mix of both

Categories of Assessments: Listening to Students♪

Online surveys:

Even End of Term Faculty/Course Surveys



Other forms of evaluation

- Grades
- Common Exam Questions
- GRE
- Where people get jobs?
- How happy they are?